

I am interested in applied machine learning and scalable algorithms, especially where there are fundamental improvements with practical impact that can be solved at scale. I have worked on problems with approximation/sketching algorithms, streaming data, machine learning and distributed systems/NoSQL. I have been lucky enough to work in industry, startups and academia, and enjoy theory and pragmatic engineering.

Employment

2015–present	<p><i>Chief Scientist, Insidesales.com (San Mateo, CA)</i></p> <p>Lead a team applying machine learning to improve efficiency of sales teams. We process both structured CRM data and unstructured sources such as audio (sales calls), emails and external data. We use a variety of ML techniques including supervised and reinforcement learning.</p>
2014–15	<p><i>CTO, C9 Inc (San Mateo, CA) (acquired by Insidesales.com)</i></p> <p>Applying machine learning to enterprise sales pipelines in order to predict sales activity and improve forecasting. C9 was one of the first companies to apply machine learning to enterprise sales pipelines in order to compute accurate forecasts. Mostly supervised learning, time series forecasting.</p>
2013	<p><i>Founder, Featurestream.io (Cambridge, UK)</i> <i>and visiting researcher, Cambridge Computer Laboratory</i></p> <p>Featurestream.io is an API for machine learning on streaming JSON event data. Initially intended to be a random forest implementation for Spark MLlib, I implemented a distributed algorithm for incremental, distributed random forests/ETrees based on Hoeffding trees and includes machinery for automated feature engineering, implemented on Spark streaming.</p>
2009–12	<p><i>CTO, cofounder, Acunu (London, UK)</i></p> <p>Acunu is a startup developing analytics tools for big data problems. We built an engine for analytics on streaming data using Cassandra and sketching techniques, and developed the Stratified B-tree, a write-optimized versioned cache-oblivious indexing algorithms as a replacement for CoW B-trees. We started the London Big Data meetup. By 2012 we were 30 employees (15 PhDs), 15 patents, and over \$10m in VC funding.</p>
2008–13	<p><i>Junior Research Fellow in Computer Science, Oxford, UK & fellow at St. John's College</i></p> <p>Elected by open competition. Research interests: efficient algorithms for data mining via approximation, streaming, machine learning. Took a sabbatical in 2009 to found Acunu.</p>
2006–7	<p><i>Technicolor Research, Paris</i> <i>Microsoft Research, Cambridge</i></p> <p>We used randomized bittorrent-like epidemic models to obtain provably optimal randomized algorithms for broadcasting and live streaming algorithms.</p>

Education

2013	DPhil, Oxford University (by incorporation)
2006	PhD Computer Science, King's College, Cambridge University Thesis: <i>Forbidden-Set Compact Routing</i> Introduced the forbidden-set distance problem, which deals with efficiently (with low space) precomputing approximate distances around a set of unknown failures in a graph. We use this to develop a low-stretch failure-tolerant compact routing protocol. This has proved to be a deep topic, and has been used in solutions to deeper problems such as dynamic graph connectivity. Nominated for British Computer Society Best Dissertation Award.
1999–2002	BSc Computer Science, Warwick University, UK (1st class) Thesis: <i>Lightweight web prediction algorithms</i> Best overall student each year and top graduating student in CS

Teaching

2008–11	Oxford University <i>Randomized Algorithms, Advanced Data Structures and Algorithms</i>
2003–06	Cambridge University <i>Algorithms and Data Structures, Mathematical Methods in Computer Science, Probability, Complexity Theory</i>

Interests

Sports (golf, rowing, cycling), music (played drums semi-professionally, mostly big band, jazz), cars and engineering. In 2005 I rowed for Cambridge Lightweights (CULRC Granta) against Oxford, 2002-7 King's College men's 1st VIII, vice-captain 2004. Also various invited talks, program committee work.

Selected Publications

- 2017 A. Twigg. Persistent Cache-oblivious Streaming Indexes. *ArXiv*, abs/1707.08186, July 2017
- 2015 Andrew Twigg and Eduardo C. Xavier. Locality-preserving allocations problems and coloured bin packing. *Theoretical Computer Science*, 596:12 – 22, 2015
- 2011 Andrew Twigg, Andrew Bye, Grzegorz Miłos, Tim Moreton, John Wilkes, and Tom Wilkie. Stratified B-trees and versioned dictionaries. HotStorage'11, pages 10–10, Berkeley, CA, USA, 2011. USENIX Association
- Andrew Bye and Andy Twigg. Optimal query/update tradeoffs in versioned dictionaries. *ArXiv*, abs/1103.2566, 2011
- 2010 Bruno Courcelle and Andrew Twigg. Constrained-path labellings on graphs of bounded clique-width. *Theory Comput. Syst.*, 47(2):531–567, 2010
- 2008 Thomas Bonald, Laurent Massoulié, Fabien Mathieu, Diego Perino, and Andrew Twigg. Epidemic live streaming: optimal performance trade-offs. In *SIGMETRICS*, pages 325–336, 2008
- Andrew Twigg. Worst-case time decremental connectivity and k-edge witness. *ArXiv*, abs/0810.5477, 2008
- Bruno Courcelle, Cyril Gavoille, Mamadou Moustapha Kanté, and Andrew Twigg. Connectivity check in 3-connected planar graphs with obstacles. *Electronic Notes in Discrete Mathematics*, 31:151–155, 2008
- Laurent Massoulié and Andrew Twigg. Rate-optimal schemes for peer-to-peer live streaming. *J. Performance Evaluation*, 65(11-12):804–822, 2008
- Rahul Sami and Andy Twigg. Lower bounds for distributed markov chain problems. *ArXiv*, abs/0810.5263, 2008
- 2007 Laurent Massoulié, Andrew Twigg, Christos Gkantsidis, and Pablo Rodriguez. Randomized decentralized broadcasting algorithms. In *INFOCOM*, pages 1073–1081, 2007
- Bruno Courcelle, Cyril Gavoille, Mustapha Kante, and Andrew Twigg. Forbidden-set labelling on graphs. *PODC workshop on Locality Preserving Distributed Computing Methods (LOCALITY)*, 2007
- Bruno Courcelle and Andrew Twigg. Compact forbidden-set routing. In *STACS 2007, 24th Annual Symposium on Theoretical Aspects of Computer Science, Aachen, Germany*, pages 37–48, 2007
- Karl Krukow and Andrew Twigg. The complexity of fixed point models of trust in distributed networks. *Theoretical Computer Science*, 389(3):528–549, 2007
- 2006 Andrew D. Twigg. Compact forbidden-set routing (PhD Thesis). Technical Report UCAM-CL-TR-678, University of Cambridge, Computer Laboratory, December 2006
- Laurent Massoulié, Andrew Twigg, Christos Gkantsidis, and Pablo Rodriguez. Provably optimal decentralized broadcasting algorithms. Technical report, 2006. MSR-TR-2006-105
- 2005 Karl Krukow and Andrew Twigg. Distributed approximation of fixed-points in trust structures. In *ICDCS*, pages 805–814, 2005
- 2003 Andrew Twigg and Nathan Dimmock. Attack-resistance of computational trust models. In *WETICE*, pages 275–280, 2003
- Tim Moreton and Andrew Twigg. Trading in trust, tokens, and stamps. In *Workshop on Economics of Peer-to-Peer Systems*, 2003
- Andrew Twigg. A subjective approach to routing in p2p and ad hoc networks. In *iTrust*, pages 225–238, 2003
- Tim D. Moreton and Andrew Twigg. Enforcing collaboration in peer-to-peer routing services. In *iTrust*, pages 255–270, 2003